AD_____

AWARD NUMBER: W81XWH-07-1-0545

TITLE: Highly Conserved, Large, Non-Coding Transcripts and their Role in the Development of Breast Cancer

PRINCIPAL INVESTIGATOR: David I. Smith, Ph.D.

CONTRACTING ORGANIZATION: Mayo Clinic
Rochester, MN 55905

REPORT DATE: August 2009

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
31-AUG-2009	Final	8/1/2007-7/31/2009

4. TITLE AND SUBTITLE
Highly conserved, large, non-coding transcripts and their role in the development of breast cancer

5a. CONTRACT NUMBER
W81XWH-07-1-0545

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
David I Smith, PhD
Email: smith.david@mayo.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
Mayo Clinic
200 1st Street SW
Rochester, MN 55905

8. PERFORMING ORGANIZATION REPORT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research & Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Non-coding transcripts, Stress-induction, Cellular Regulators

14. ABSTRACT

The sequencing of the human genome has revealed that only 2% of the genome actually codes for protein. However, the remainder of the genome is not "junk" and it has recently been revealed that most of the genome is transcriptionally active. We utilized a tiling array approach to examine the entire genome for transcriptional activity and found a large number of non-coding transcripts. When we originally proposed the work in this Concept Award, we proposed to study a group of long, highly conserved, non-coding transcripts which had altered expression and sometimes mutations in breast cancer. We have subsequently found that many of these highly conserved transcripts are either part of known genes or highly homologous to known genes. However, we've now identified two new groups of novel non-coding transcripts which are not part of genes. The first group is non-coding transcripts that have increased expression in response to the DNA damage induced by the carcinogen NNK. The second group is identified by analyzing breast cancer cell lines with tiling arrays searching for non-coding transcripts that had consistently altered expression.

We chose one of the NNK-induced transcripts, NIT 5, and one of the breast cancer altered transcripts, bcNCT 28, and began to characterize them further. Here we describe our final report on the work done with these two transcripts to determine their role in breast cancer.

15. SUBJECT TERMS : None provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	16	**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

ABSTRACT

The sequencing of the human genome has revealed that only 2% of the genome actually codes for protein. However, the remainder of the genome is not “junk” and it has recently been revealed that most of the genome is transcriptionally active. We utilized a tiling array approach to examine the entire genome for transcriptional activity and found a large number of non-coding transcripts. When we originally proposed the work in this Concept Award, we proposed to study a group of long, highly conserved, non-coding transcripts which had altered expression and sometimes mutations in breast cancer. We have subsequently found that many of these highly conserved transcripts are either part of known genes or highly homologous to known genes. However, we’ve now identified two new groups of novel non-coding transcripts which are not part of genes. The first group is non-coding transcripts that have increased expression in response to the DNA damage induced by the carcinogen NNK. These NNK-induced transcripts (NiTs) are all over 300 nucleotides long and have altered expression in breast cancer. We have validated these transcripts and have utilized Northern blots to determine the precise size of these transcripts (and they are between 500 and 1,500 base pairs in length). The second group we identified by analyzing breast cancer cell lines with tiling arrays searching for non-coding transcripts that had consistently altered expression. We’ve now identified a group of breast cancer non-coding transcripts (bcNCTs). These have been validated in a larger panel of breast cancer cell lines. In this final report for this Concept Award, we describe our selection of one of the NiTs (NiT 5) and one of the bcNCTs (bcNCT 28), as well as our preliminary characterization of these two long non-coding transcripts. We also describe our work to determine the role that these long non-coding transcripts play in the development of breast cancer.

INTRODUCTION

Only 2% of the human genome actually codes for protein. In spite of this, it turns out that most of the genome is transcriptionally active. The question we set out to answer is what the function of these non-coding transcripts are and what role, if any, do they play in the development of breast cancer. Our major hypothesis is that these non-coding transcripts play an important regulatory role within the cells and would be important targets of alteration during the development of breast cancer.

We decided to focus our efforts on a group of long (greater than 300 nts) non-coding transcripts that we identified by using a tiling array approach to identify transcriptional units across the genome. We specifically focused on the most highly conserved of these non-coding transcripts and had preliminary evidence that several of these were also targets of alteration (both in terms of expression and as occasional mutational targets) in breast cancer. We proposed to characterize two of these (NiT4 and NiT5) as they were the most highly conserved.

However, subsequent analysis of the highly conserved non-coding transcripts has revealed that most of the more highly conserved transcripts actually had homology to existing coding transcripts. Indeed results of the ENCODE project now reveal that each coding gene produces on the average about 5 distinct transcripts and that there is much greater complexity to gene organization than previously anticipated. In addition, the simple model of genes being merely a collection of contiguous exons that are spliced together may also be wrong. Many transcripts generated are actually produced from quite disparate chromosomal regions (this was revealed most convincingly by doing mate-pair sequence analysis of large numbers of transcripts using Next Generation DNA sequencing technology). In addition, there are cryptic exons that are sometimes hundreds of kilobases upstream or downstream from the simple organized gene. Indeed, many of the highly conserved non-coding transcripts that we were characterizing were found to be linked to known genes either due to extensive homology to those known genes or because they actually corresponded to those distant exons. However, the group of Dr. Carlo Croce recently described the identification of a number of highly conserved long non-coding transcripts which appear to be bona-fide non-coding transcripts that are altered in human leukemias and carcinomas (Calin GA et al: Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 2007; 12: 215-229).Fortunately, we had two additional sources of non-coding transcripts that we began to analyze in greater detail.

Research Accomplishment 1: Identify NNK-induced long non-coding transcripts (NiTs)
We performed our initial tiling array experiment to identify possible non-coding transcripts, we examined a normal human bronchial epithelial cell line exposed to the DNA damaging tobacco carcinogen NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone).

We analyzed the entire genome in response to NNK stress looking for long potential non-coding transcripts (longer than 300 base pairs) that were either induced or repressed by this stress. This analysis identified transcripts that were induced by exposure to NNK (called NiTs for NNK-induced non-coding transcripts) and suppressed by exposure to NNK (called NsTs for NNK-suppressed non-coding transcripts; Figures 2 and 3, respectfully).

We continued to focus our attention on longer non-coding transcripts as we attempted to look for novel transcripts that were not either miRNAs or potential miRNA precursors. It is interesting and important to note that two previously reported non-coding transcripts, tncRNA and MALAT-1 which are located adjacent to each other on chromosome 11, were identified in this screen as being induced by exposure to NNK and thus qualify as NiTs. Figure 1 below shows the integrated genome browser (IGB) results with tncRNA and MALAT-1, as well as several of the newly identified stress-responsive non-coding transcripts.

Figure 1: NNK induction of long non-coding RNAs tncRNA and MALAT

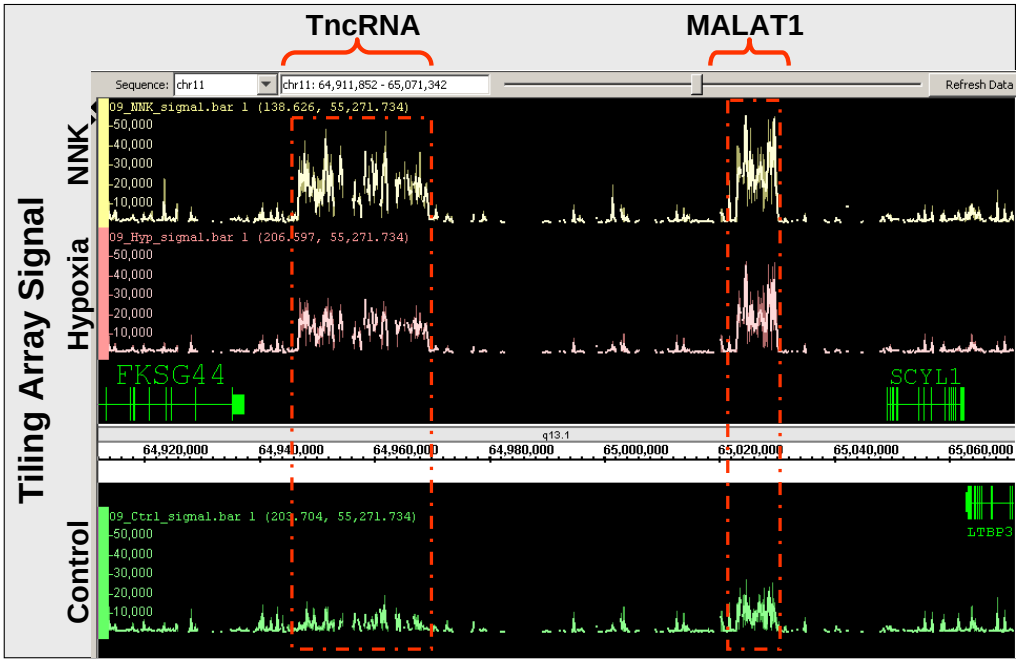


Figure 2: NNK-induced non-coding transcript (NCT)

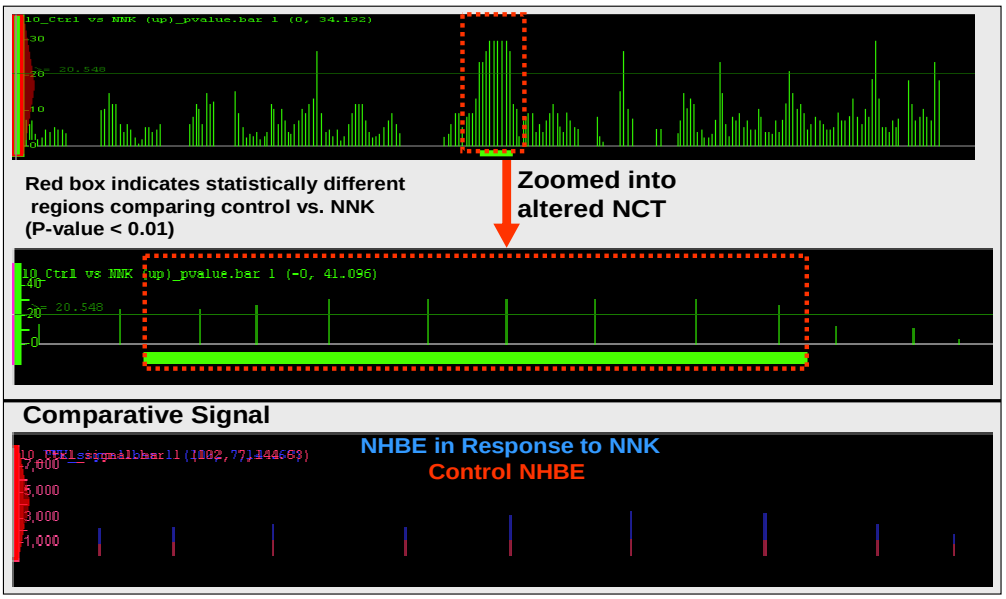
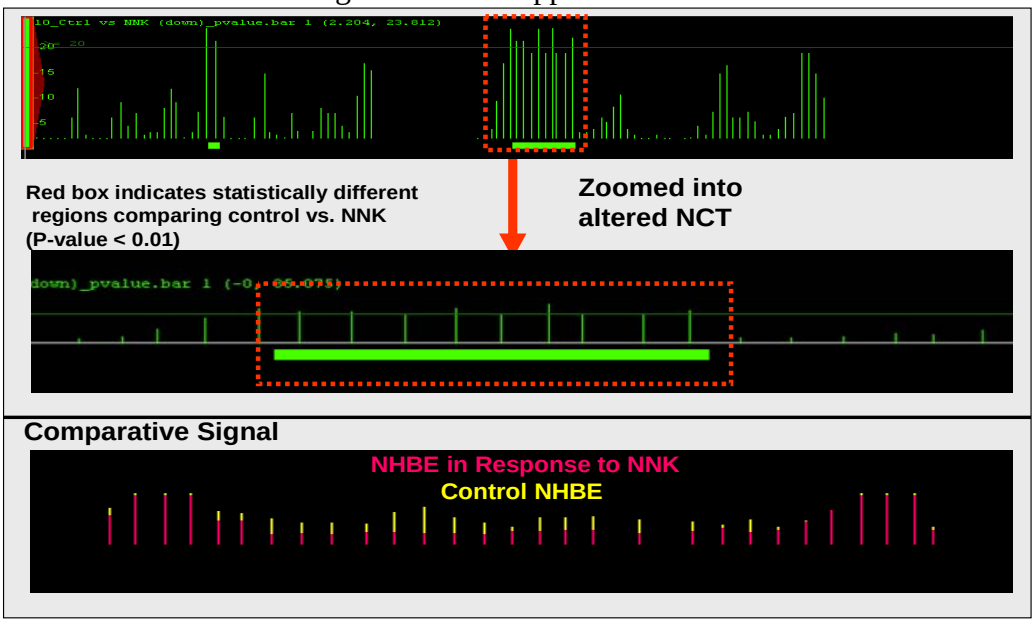


Figure 3:NNK-suppressed NCT



We next began to more fully characterize the NiTs; we have not yet begun to work on the NSTs. We identified over 231, 305 NNK-Transcriptionally active regions (TARs) within the entire human genome with their lengths ranging from 100 nt to greater than 3000 nt. Of these TARS 119,305 were induced by NNK, while 112,000 were decreased by NNK treatment. Currently, we have analyzed the NNK-induced TARS further by identifying the longer non-coding transcripts and by excluding transcripts smaller than 300 nts, thereby identifying 1,305 long NNK-induced TARs (Figure 4). We then used stringent criteria to identify true long NNK-induced non-coding transcripts. This criteria includes non-coding potential determined by verifying a lack of significant open reading frames, the NiTs also had no homology to any RefSeq Genes or any potential ncRNAs (C/D and H/ACA, snoRNAs and microRNAs), and contain a lack of homology to any known mRNAs. In addition, the criteria demanded that NiTs not have duplications or repeats. The intense final analysis of the NNK-TARs allowed us to identify 12 NiTs (Figure 5) excluding tncRNA and MALAT1.

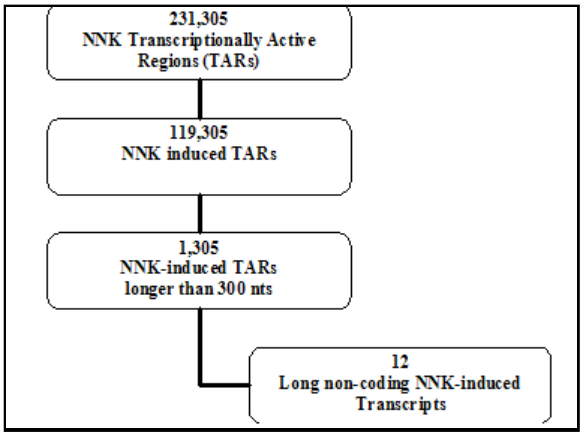
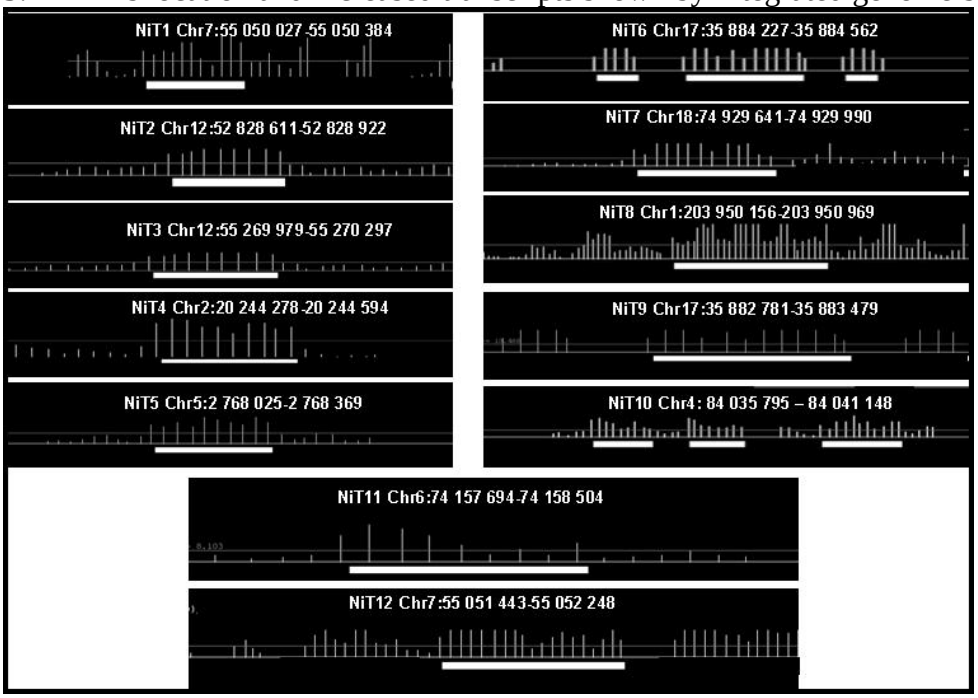


Figure 4: Identification of NiTs (NNK-induced non-coding Transcripts)

We also generated probes for these stress-responsive non-coding transcripts and hybridized them to Northern Blots to determine the size of the putative non-coding transcripts. Figure shows several representative NiTs (NiT2 and NiT6) Northern blots. This analysis revealed that the full size of these transcripts was greater than that determined on the tiling arrays.

Figure 5: 12 NiTs location and increased transcripts shown by Integrated genome browser



Research Accomplishment 2: Validate NiT expression levels using quantitative real-time PCR (qPCR) and Northern blot analysis

Specific primers were determined for optimal concentrations for quantitative real-time PCR (qPCR). Once this was accomplished we began to analyze panels of random-primer cDNAs made from different RNA samples. This included a panel of RNAs isolated from various normal human tissues (Figure 6), a panel of breast normal and cancer-derived cell lines (Figure 7 and Table 1), as well as normal human bronchial epithelial (NHBE) and MCF7 cell lines exposed to NNK (Figure 8). The goal of this was to determine whether these transcripts were indeed stress responsive as we had observed in the tiling array experiment and validate those results, examine the spectrum of their expression in different tissues, and finally whether or not they have altered expression in both breast cancer cell lines and primary tumors.

Figure 6: NiTs expression in normal human tissue

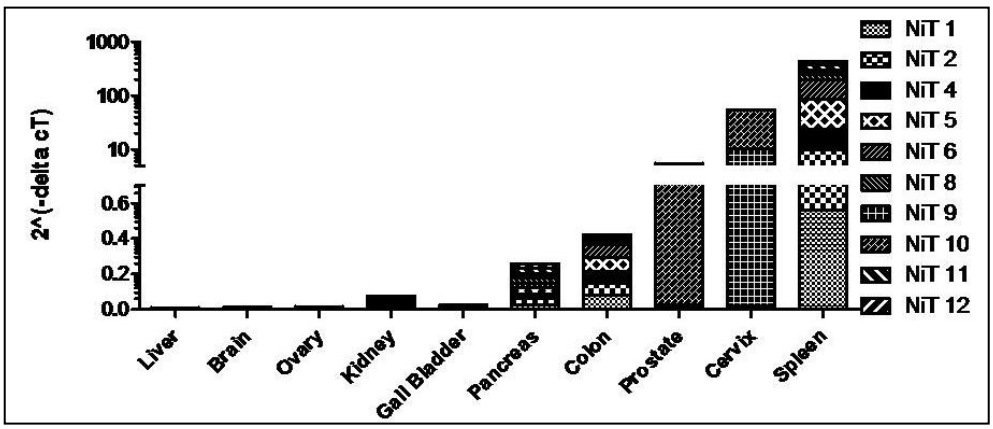


Figure 7: NiTs expression in breast cancer panel

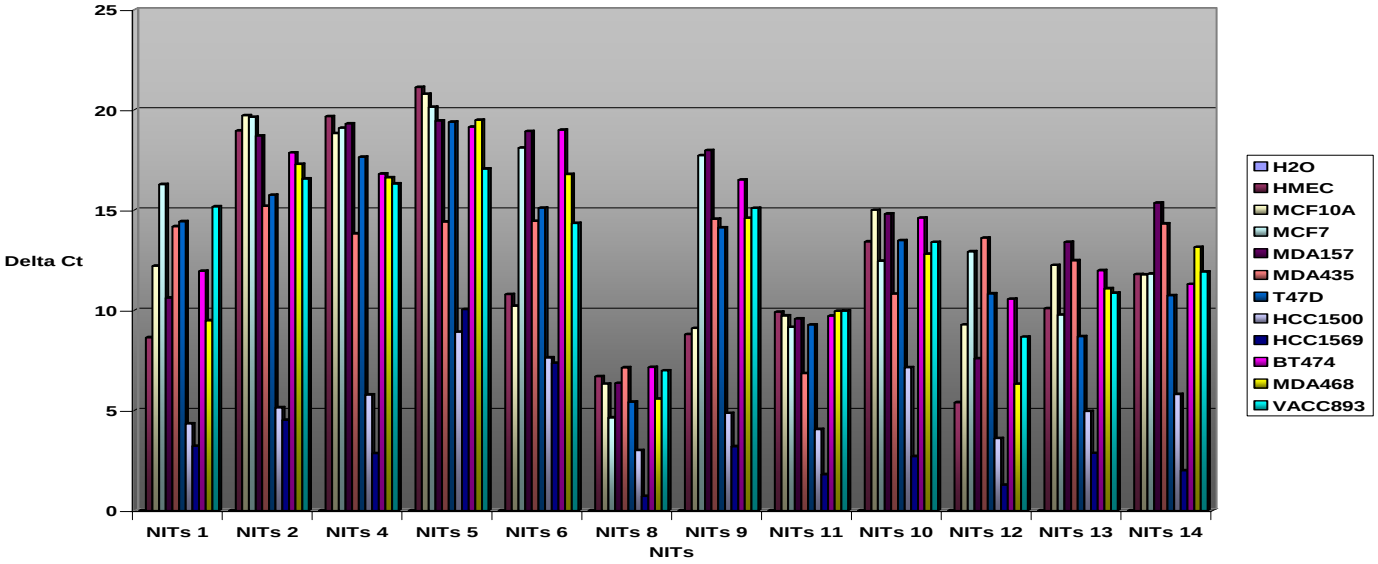


*Table 1: NiT expression indicated by average delta C_T relative to actin control transcript with standard deviation in a breast cancer cell line panel. Significance listed with * indicated by 2-fold difference compared to normal HMEC*

	NiT1	NiT2	NiT4	NiT5	NiT6	NiT8	NiT9	NiT10	NiT11	NiT12
HMEC	8.6±1.01	18.9±1.09	19.6±0.00	21.1±0.15	10.8±0.16	6.7±0.02	8.8±0.21	13.4±0.19	9.9±0.03	5.4±0.05
MCF10A	12.2±0.23	19.7±0.09	18.8±0.12	20.8±0.00	10.2±2.33	6.3±0.03	9.1±0.14	15.0±0.07	9.7±0.02	9.3±0.16
MCF7	16.3±0.03	17.4±0.73*	16.8±0.86*	18.1±0.34*	18.1±0.21	4.6±0.18	17.7±0.07	12.8±0.15	9.2±0.25	12.9±0.14
MDA157	10.6±0.43	18.7±0.84	19.3±0.21	19.4±0.00*	18.9±0.75	6.4±0.01	18.0±1.26	14.8±0.31	9.6±0.16	7.6±0.04
MDA435	14.2±0.18	15.2±0.63*	13.8±0.05*	14.4±0.44*	14.4±0.56	7.1±0.17	14.5±0.10	10.8±0.29*	6.8±0.36*	13.6±0.35
T47D	14.4±0.06	14.2±0.07*	16.2±0.47*	18.8±0.00*	15.1±0.48	5.4±0.19*	14.1±0.15	13.9±0.10	9.3±0.18	10.8±0.27
HCC1500	4.3±0.17*	5.1±0.01*	5.8±0.06*	8.9±3.57*	7.6±1.01*	3.0±0.15*	4.9±0.38*	7.1±0.16*	4.1±0.08*	3.6±0.27*
HCC1569	3.2±2.68*	4.5±1.93*	2.9±3.01*	10.0±1.00*	7.4±1.00*	0.7±1.54*	3.2±0.09*	2.7±1.24*	1.8±2.57*	1.3±0.01*
BT474	11.9±5.13	17.9±1.00*	16.8±1.00*	19.1±1.00*	19.0±1.00	7.2±0.05	16.5±0.45	14.6±0.21	9.7±0.40	10.6±0.16
MDA468	9.5±1.47	14.8±0.46*	16.6±0.00*	19.4±1.23*	16.8±2.32	5.6±0.02*	14.6±0.74	12.8±0.05*	10.0±0.02	6.3±0.00
UACC893	15.2±0.06	16.6±0.08*	16.3±0.00*	17.1±1.89*	14.3±0.02	7.0±0.19	15.1±0.48	13.4±0.02	10.0±0.00	8.7±0.05

Figure 8: NiTs expression in response to NNK treatment

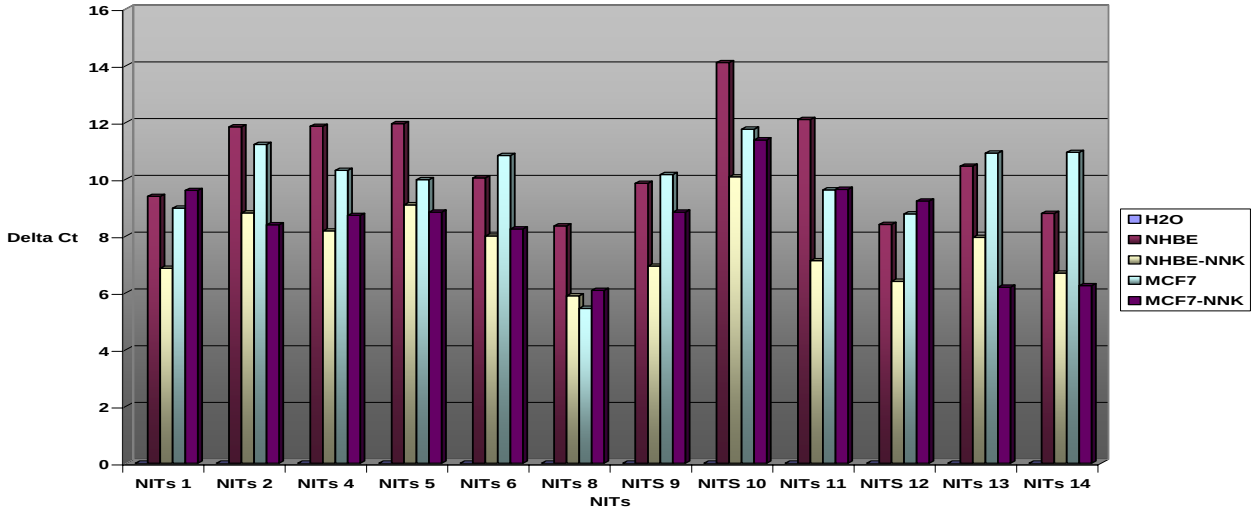
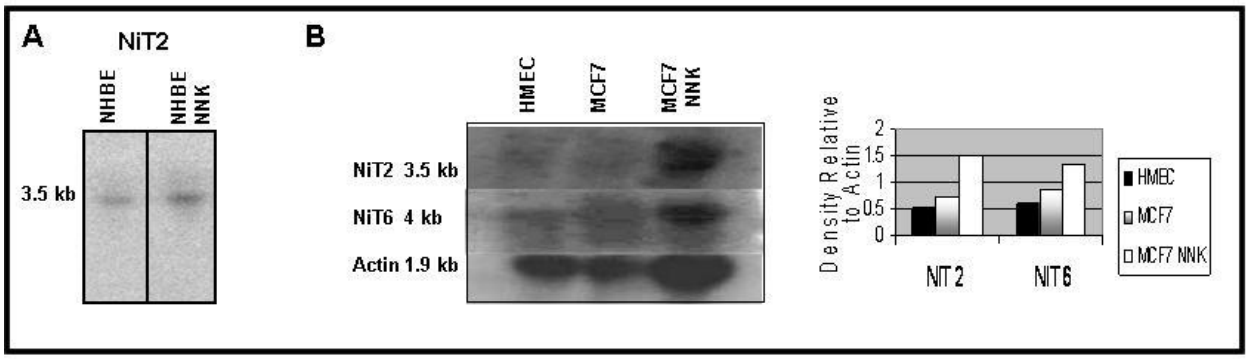


Figure 9: NiT2 and NiT6 RNA transcript in NHBE and breast cell lines



Of the 12 NiTs that have been characterizing, we decided to focus all of our efforts on one particular NiT, NiT 5. This NiT had increased expression in the majority of the breast cancer cell lines, as well as primary tumors, that we been characterized. NiT5 was identified on chromosome 5:2 768 026 – 2 768 369, it is intergenically located and is several megabase to kilobases away from any nearby genes. NiT5 has high expression in normal proliferative tissues such as colon and spleen and contains low expression in slower proliferative normal tissues such as brain (Figure 10). NiT5 also has higher expression in panel of breast, ovarian, and cervical cancer cell lines, however has lower expression in an endometrial primary tissue panel (Figures 11 and 12).

Figure 10: NiT5 expression in normal tissue panel

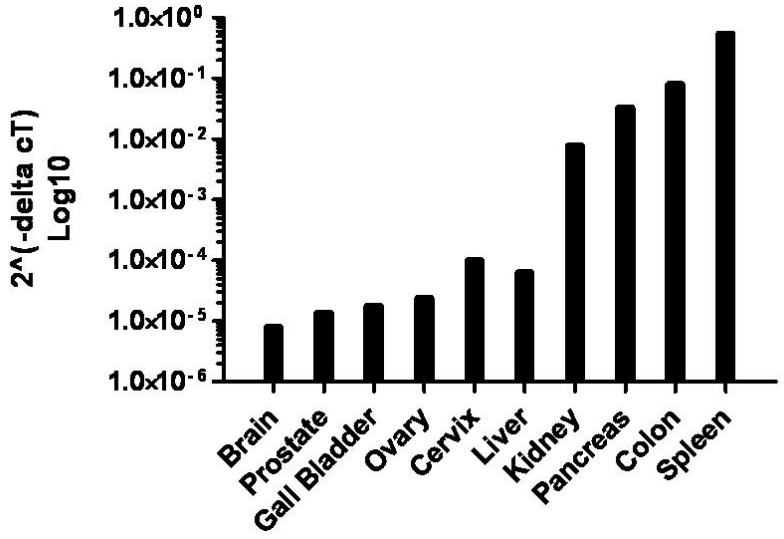


Figure 11: NiT5 has high expression in breast, ovarian, and cervical cancer cell line panels

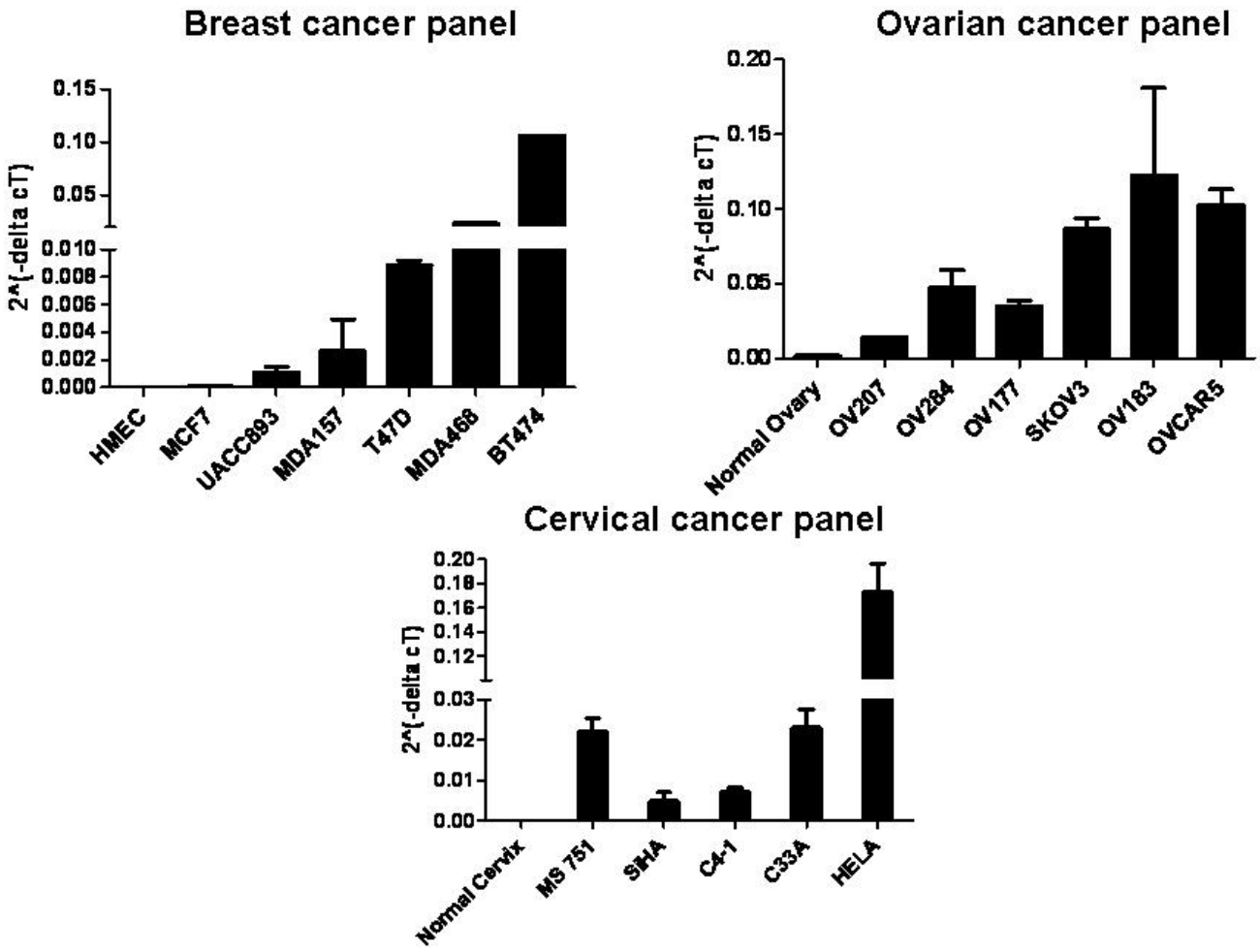
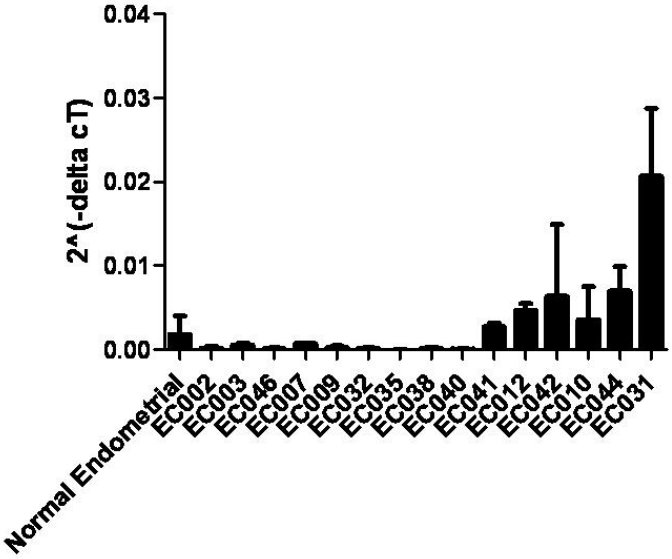


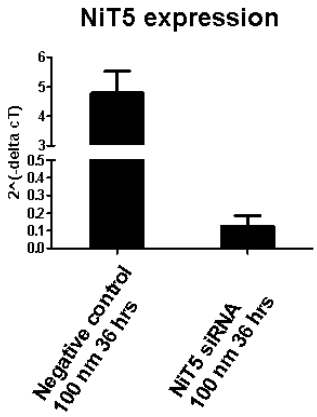
Figure 12: NiT5 has low expression in endometrial primary tissue panel



Research Accomplishment 3: Attempts to knock down expression of NiT5 transcript using siRNA technology.

The last part of our proposed work on NiT 5 was to determine what effect would be observed in cells after we had knocked down the expression of NiT 5 (using siRNA) in cell lines that had increased expression of NiT 5. We chose two cell lines **BT474** and **OvCar 5** (the first is a breast cancer cell line and the second is an ovarian cancer cell line). We constructed several siRNAs to attempt to knock down the expression of NiT 5 but none of the siRNA constructs tested produced a significant decrease in NiT 5 expression. We tried several different approaches to successfully transfect functional siRNAs into the cell lines and then realized that perhaps the problem was that we were dealing with a non-coding transcript which is localized in the nucleus and all our various approaches to get the siRNAs into the cells would localize them into the cytoplasm. We then searched for protocols that would transfect nuclear encoded RNAs and eventually found one that does result in decreased NiT 5 expression. Figure 13 summarizes the level of knock-down of NiT 5 in **BT474** and demonstrates that we can indeed decrease the nuclear expression of NiT 5. Unfortunately, these results were just obtained several weeks ago (actually after the termination date on this grant). As a result of this we do not yet have any results on the phenotypic effect of this.

Figure 13: siRNA knockdown of NiT5 expression in BT474



In the progress report from the last year we discussed performing knock-downs of both NiT 5 and NiT 4, but because of the difficulty that we had in knocking down the expression of nuclear-localized non-coding transcripts, we have not yet begun to work on decreasing the expression of NiT 4.

Research Accomplishment 4: Identification of long non-coding transcripts altered in breast cancer (bcNCTs)

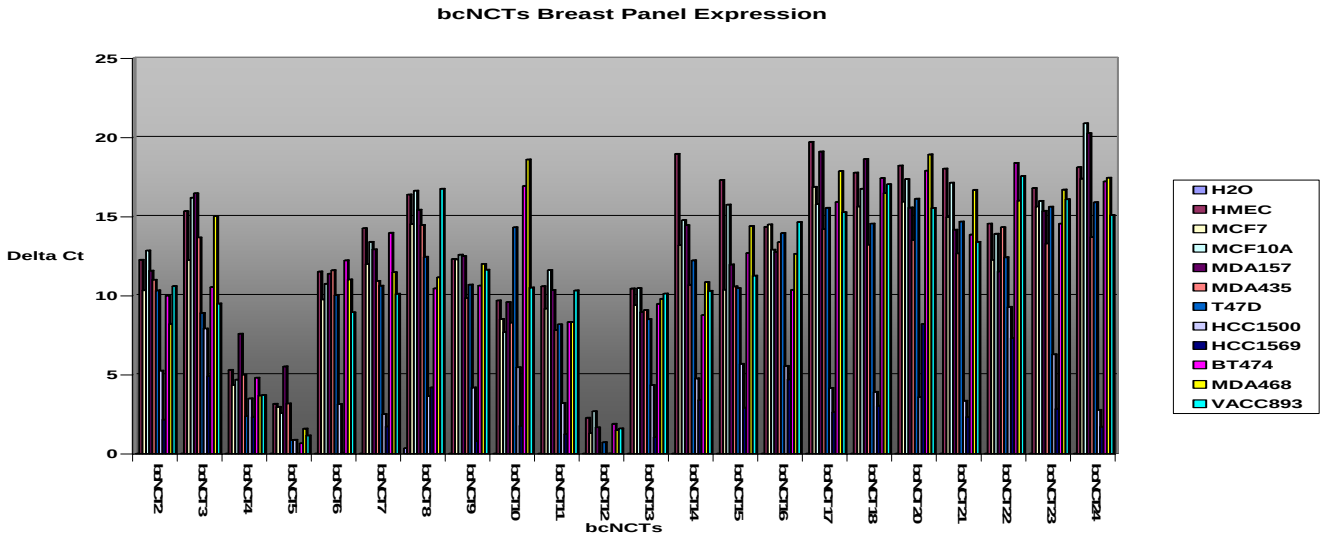
The second source of potential non-coding transcripts we are studying were derived from an experiment that we conducted to directly search for non-coding transcripts that were altered in breast cancer cell lines. We took the normal breast epithelial cell line HMEC and five breast cancer cell lines including HCC1500, HCC1569, MDA157, T47D, and MDA435. Total RNA was isolated from each of these and then hybridized to a single tiling array chip in triplicate. While our initial studies were completed with 14 tiling array chips covering the entire genome, we decided to conduct a more focused and considerably less expensive experiment of only examining a single tiling array chip containing probes covering three human chromosomes (chromosomes 8, 11, and 12). We then searched for non-coding transcripts with altered expression in more than one breast cancer cell line, relative to HMEC. While there were many transcripts altered in just one of the cell lines, common alterations in multiple cells lines were few. There were 24 transcripts altered in two cell lines, and 1 transcript altered in three cell lines. No transcripts were found to be altered in all four cell lines and these transcripts had

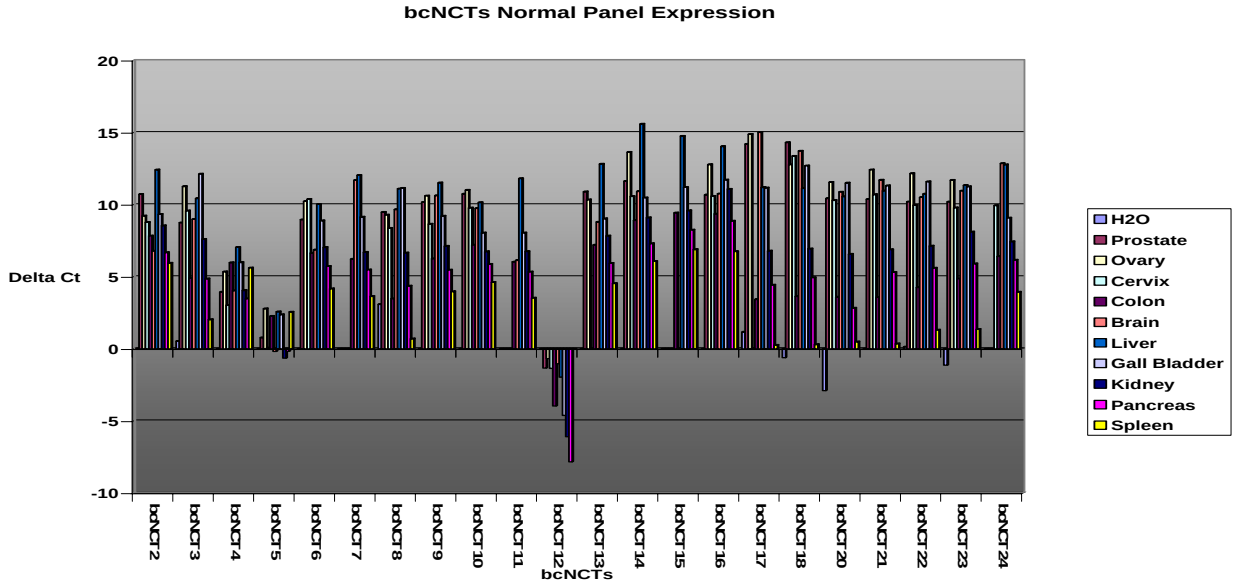
no correlation with NNK responsive NCTs. The transcripts that had these consistent alterations were called bcNCT (for breast cancer Non-Coding Transcripts).

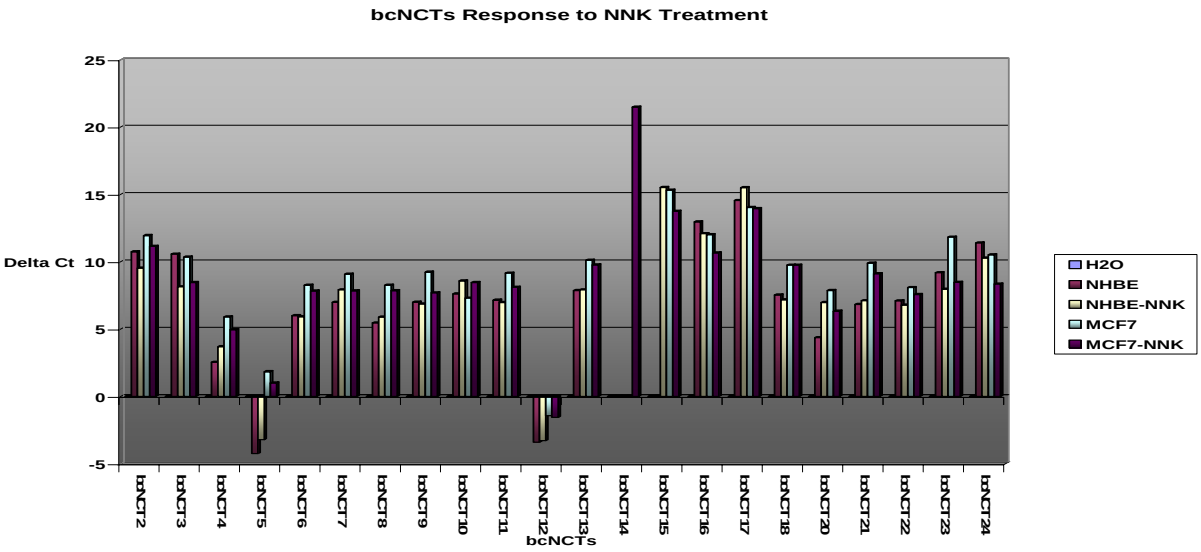
Research Accomplishment 5: Validate expression levels of bcNCT transcripts using quantitative real-time PCR (qPCR)

The 24 bcNCTs altered in the two cell lines were then characterized exactly as the NiTs were characterized. Real-time primers were designed and optomized to verify whether or not the bcNCT expression was indeed consistently altered in a much larger panel of breast cancer cell lines, as well as primary tumors. We also characterized the bcNCTs to determine expression in normal human tissues. The bcNCTs were also characterized to identify if they were stress responsive, as this was our original goal with the stress response tiling array experiment. It is important to note there that the two characterized non-coding transcripts, tncRNA and MALAT-1, which have increased expression in response to DNA damage induced by NNK also have increased expression in several of the breast cancer cell lines. This provides us good support for our hypothesis that stress-responsive non-coding transcripts are good candidates for transcripts that also have altered expression during the development of breast cancer. The Figures below show several of the bcNCTs that we have chosen for further study.

Figure 14: Expression Patterns of bcNCT transcripts


bcNCTs Breast Panel Expression
Delta Ct
25
20
15
10
5
0
bcNCT2
bcNCT3
bcNCT4
bcNCT5
bcNCT6
bcNCT7
bcNCT8
bcNCT9
bcNCT10
bcNCT11
bcNCT12
bcNCT13
bcNCT14
bcNCT15
bcNCT16
bcNCT17
bcNCT18
bcNCT20
bcNCT21
bcNCT22
bcNCT23
bcNCT24
bcNCTs
H2O
HMEC
MCF7
MCF10A
MDA157
MDA435
T47D
HCC1500
HCC1569
BT474
MDA468
VACC893


bcNCTs Normal Panel Expression
Delta Ct
20
15
10
5
0
-5
-10
bcNCT2
bcNCT3
bcNCT4
bcNCT5
bcNCT6
bcNCT7
bcNCT8
bcNCT9
bcNCT10
bcNCT11
bcNCT12
bcNCT13
bcNCT14
bcNCT15
bcNCT16
bcNCT17
bcNCT18
bcNCT20
bcNCT21
bcNCT22
bcNCT23
bcNCT24
bcNCTs
H2O
Prostate
Ovary
Cervix
Colon
Brain
Liver
Gall Bladder
Kidney
Pancreas
Spleen



We also previously discussed our plans to decrease the expression of bcNCT 28, but because of the difficulty that we had in knocking down NiT 5 expression, we did not have time to get to examine what effect alterations in bcNCT 28 would have on different cell lines.

A second experiment that we originally proposed, **but did not have sufficient time to conduct**, was to test the effect of increasing the expression of NiT 4, NiT 5, and bcNCT 28 in cell lines derived from normal bronchial epithelial cells (NHBE). Our hypothesis was that this would result in, at a minimum, increased cellular growth and potentially more profound phenotypic changes. These experiments are currently underway for the NiT5 construct. Currently, over-expression constructs are being created by cloning full length NiT5 into a TOPO dual promoter SP6/T7 and creating plasmid constructs for transfection into NHBE cell lines. These constructs will be transfected using RNAifect reagents and the resulting transformants will be used to conduct functional studies of migration and proliferation.

It should be mentioned that thanks to the Department of Defense Breast Cancer Research Program we have just received a Breast Cancer Idea award to continue our studies on NiT 5. We are currently performing these experiments and will then be able to determine what effect this has on both cancer-derived cell lines with increased expression of these non-coding transcripts as well as increasing non-coding transcript expression in normal tissue-derived cell lines.

Another set of experiments that we proposed to conduct, but have not yet completed because of our difficulty in the knock-down experiments, was to utilize the powerful technology of Next Generation RNA sequencing to examine the genes that have altered expression after either increasing or decreasing the expression of NiT 5. However, these experiments will be conducted as part of our Breast Cancer Idea award. We are currently concluding a set of successful experiments to optimize the siRNA transfection system for knocking down expression of NiT5 transcripts in BT474 cells in culture. We will begin the sequencing of RNA isolated from BT474 cells and BT474 cells with siRNA reduced expression of NiT5 using whole transcriptome sequencing technology on the Illumina platform. Pathway analysis will be performed to identify transcripts, genes and pathways that are differentially regulated following silencing of NiT5.

We would like to thank the Department of Defense Breast Cancer Program for supporting our work. While it is still in the preliminary stages, we remain hopeful that this work will identify important new targets of alteration during the development of breast cancer. We have submitted an RO1 proposal to the NIH on some of the non-

coding transcripts, but as expected, the first submission was triaged out. However, with additional supportive results (which we hope to obtain with Department of Defense support), this grant should be more competitive.

KEY RESEARCH ACCOMPLISHMENTS

(1) Identified NNK-induced long non-coding transcripts (NITs).
(2) Validated these NIT transcripts both with real-time RT-PCR and with Northern blot analysis.
(3) Attempted to knock down the expression of NiT 5 using siRNA technology.
(4) Identified breast cancer altered long non-coding transcripts (bcNCTs).
(5) Validated these bcNCT transcripts both with real-time RT-PCR and with Northern blots.

PUBLICATIONS AS A DIRECT RESULT OF THIS GRANT

Perez DS, Hoage TR, Pritchett JR, Ducharme-Smith AL, Halling ML, Ganapathiraju SC, Streng PS, Smith DI. Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Molec Genet* 2008; 17: 642-655.

Silva J, Perez DS, Ducharme-Smith AL, Pritchett JR, Smith DI. NNK-induced transcripts: A new group of large stress responsive non-coding transcripts have altered expression in breast cancers. *In Press, Genomics,* 2009.

ABSTRACTS AS A RESULT OF THIS GRANT

Silva JA, Perez DS, Smith DI. A new class of novel long NNK-induced non-coding transcripts. Abstract #08-AB-2956- AACR, 2008 Annual AACR Meeting

Silva JA, Pritchett JR, Halling MR, Perez DS, Smith DI. Long NNK-induced non-coding transcripts have altered expression in breast cancer cells. Abstract #09-AB-579, 2009 Annual AACR Meeting.

Ma X-H, Pritchett JR, Halling M, Smith DI. Involvement of long non-coding RNAs in human breast cancer development. Abstract #09-AB-2956, 2009 Annual AACR Meeting.